

# Prediction of Collision Cross-Section Values for Extractables and Leachables from Plastic Products

Xue-Chao Song, Nicola Dreolin, Elena Canellas, Jeff Goshawk, and Cristina Nerin\*



Cite This: *Environ. Sci. Technol.* 2022, 56, 9463–9473



Read Online

ACCESS |



Metrics & More



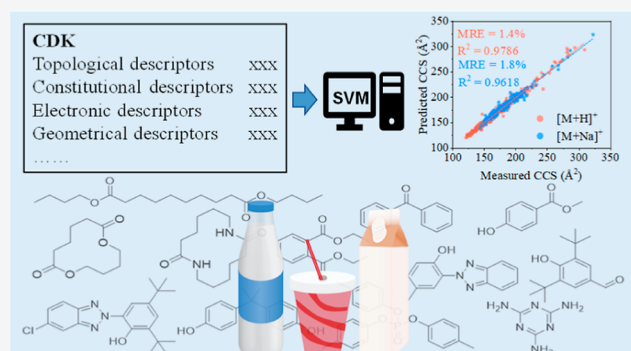
Article Recommendations



Supporting Information

**ABSTRACT:** The use of ion mobility separation (IMS) in conjunction with high-resolution mass spectrometry has proved to be a reliable and useful technique for the characterization of small molecules from plastic products. Collision cross-section (CCS) values derived from IMS can be used as a structural descriptor to aid compound identification. One limitation of the application of IMS to the identification of chemicals from plastics is the lack of published empirical CCS values. As such, machine learning techniques can provide an alternative approach by generating predicted CCS values. Herein, experimental CCS values for over a thousand chemicals associated with plastics were collected from the literature and used to develop an accurate CCS prediction model for extractables and leachables from plastic products. The effect of different molecular descriptors and machine learning algorithms on the model performance were assessed. A support vector machine (SVM) model, based on Chemistry Development Kit (CDK) descriptors, provided the most accurate prediction with 93.3% of CCS values for  $[M + H]^+$  adducts and 95.0% of CCS values for  $[M + Na]^+$  adducts in testing sets predicted with <5% error. Median relative errors for the CCS values of the  $[M + H]^+$  and  $[M + Na]^+$  adducts were 1.42 and 1.76%, respectively. Subsequently, CCS values for the compounds in the Chemicals associated with Plastic Packaging Database and the Food Contact Chemicals Database were predicted using the SVM model developed herein. These values were integrated in our structural elucidation workflow and applied to the identification of plastic-related chemicals in river water. False positives were reduced, and the identification confidence level was improved by the incorporation of predicted CCS values in the suspect screening workflow.

**KEYWORDS:** ion mobility, collision cross-section, plastic products, extractables, leachables, machine learning



## 1. INTRODUCTION

Plastics play an important role in our daily life as they are used in a variety of materials, including packaging, building and construction materials, transportation, and electrical and electronic components.<sup>1</sup> It has been reported that up to 2015, approximate 6300 million metric tons of plastic waste was generated of which only 9% was recycled. The remaining plastic waste was either incinerated, accumulated in landfills, or disposed of in natural environments.<sup>2</sup> The impact of plastic waste on the environment and, subsequently, human health is of great concern due to the release of microplastics<sup>3–5</sup> and low-molecular-weight (MW) chemicals.<sup>6–9</sup> During the production of plastics, a variety of additives are incorporated into the polymeric formulations to enhance favorable characteristics and extend service life. Commonly used additives include plasticizers, flame retardants, lubricants, antioxidants, and UV stabilizers.<sup>7</sup> Such additives have been detected in indoor dust,<sup>10,11</sup> airborne particulate matters,<sup>12–14</sup> waste water,<sup>15</sup> soils,<sup>16</sup> and rivers and oceans.<sup>17–19</sup> Plastic products have

become an important source of contaminants in aquatic and terrestrial environments.

In addition to the known substances included during the production of plastic materials, non-intentionally added substances (NIASs) can also occur. Typical NIAS include impurities, oligomers, and degradation products of material components.<sup>20</sup> For example, organophosphate esters can result from the oxidation of organophosphite antioxidants in plastics and have been detected in indoor dust.<sup>21,22</sup> If plastic products are made from recycled plastics, NIAS can also include contaminants resulting from the previous use of the material or from the recycling process itself.<sup>23</sup> In recent years, the presence of perfluoroalkyl substances in plastic products has also

Received: April 21, 2022

Accepted: June 14, 2022

Published: June 22, 2022



attracted the attention of food safety and environmental authorities.<sup>24–26</sup>

The complete structural elucidation of extractables and leachables from plastics is a challenging process due to the complexity of the matrix. In recent years, ion mobility separation (IMS) coupled to high-resolution mass spectrometry (HRMS) has emerged as a promising tool for analyzing complex samples.<sup>27–31</sup> IMS can separate molecules based on their shape, size, and charge.<sup>32</sup> Collision cross-section (CCS), derived from IMS, is a physicochemical property of ions and is related to the chemical structure and three-dimensional conformation of the molecules.<sup>32</sup> In addition, since CCS measurements are independent from chromatographic and mass spectrometric conditions, as well as the sample matrix,<sup>33</sup> they can be used as an additional molecular identifier to increase the specificity and identification confidence. Celma et al.<sup>34</sup> showed that CCS of imazalil was not affected by the sample matrix, whereas the retention time (RT) deviations ranged from 0.14 to 0.30 min; the consistent CCS values provided an extra point for unknown identification. In addition, incorporation the CCS values into the annotation process can help reduce false positive identifications<sup>35</sup> and enable structural isomers to be separated and identified.<sup>36,37</sup>

Experimental CCS values of reference standards are often measured in order to confirm compound identification by comparing them to CCS values of candidate compounds in qualitative analyses. Although public CCS databases of pesticides,<sup>38,39</sup> drugs,<sup>40</sup> steroids,<sup>41</sup> mycotoxins,<sup>42</sup> and chemicals in plastic food packaging<sup>43</sup> have been established, there remain many compounds that are not included in such libraries. As a matter of fact, many experimental CCS values of chemicals in plastics are not available due to the unavailability or high price of commercial standards. In this case, theoretical CCS values can be alternatives to be used for suspect and untargeted screening analysis. Several public CCS machine learning prediction tools have appeared in recent years, such as MetCCS,<sup>44</sup> AllCCS,<sup>35</sup> CCSondemand,<sup>45</sup> CCSbase,<sup>46</sup> and DeepCCS.<sup>47</sup> Some laboratories have also developed their own CCS prediction tools for specific classes of compounds, such as pesticides,<sup>38</sup> phenolics,<sup>48</sup> and drugs.<sup>49</sup> Many CCS values, belonging to different chemical classes, can provide a high structural diversity, and as such, the developed model can provide satisfactory prediction results for diverse chemical classes. At the time of writing, there are 3539, 7325, 7405, and 2439 CCS values in the data sets used by AllCCS, CCSondemand, CCSbase, and DeepCCS, respectively.

In a previous study,<sup>43</sup> 635 CCS values derived from 488 standards associated with plastic packaging were used to develop a support vector machine (SVM) model to predict CCS values. The CCS values of 92.6% of protonated molecules were predicted with an error of less than 5%. The CCS values of some halogenated compounds were inaccurately predicted due to the lack of halogenated compounds in the training set. Consequently, in this study, additional experimental CCS values of molecules related to plastics have been collected from the literature, with the aim of achieving more accurate CCS prediction for chemicals found in plastics. The effect of different molecular descriptors (MDs) and algorithms on the accuracy of the CCS prediction were also explored. Following optimization and external validation, the model was used to predict CCS values of molecules in two plastic-related databases: the Chemicals associated with Plastic Packaging Database (CPPdb)<sup>50</sup> and the Food Contact Chemicals

Database (FCCdb).<sup>51</sup> FCCdb also contains many plastic-related chemicals since approximately 37% of food contact materials (FCMs) are made from plastics.<sup>52</sup> The two databases were subsequently converted into screening libraries, containing the predicted CCS values, and used for the suspect screening of plastic-related chemicals in river water.

## 2. MATERIALS AND METHODS

**2.1. CCS Data Collection and Processing.** A total of 2145 experimental traveling wave CCS (<sup>TW</sup>CCS<sub>N<sub>2</sub></sub>) and drift tube CCS (<sup>DT</sup>CCS<sub>N<sub>2</sub></sub>) values were collected from seven recent publications,<sup>27,29,38,39,43,53,54</sup> of which 1425 and 720 CCS values were for [M + H]<sup>+</sup> and [M + Na]<sup>+</sup> ions, respectively (Table S1). The CCS values in the publication of Song and co-workers<sup>43</sup> were experimentally measured by injecting standards of chemicals associated with plastic food packaging. Four of the publications<sup>29,38,39,53</sup> include CCS values mainly for pesticides and pharmaceuticals found in environmental studies. The CCS values in these four databases were used in this study because pesticides are an important type of NIAS in plastic materials, especially those made from the recycled plastics.<sup>55</sup> Additionally, many pesticides contain halogens in their structure, as such, the predictions of CCS values for halogenated compounds will be more accurate by including the pesticides in the CCS data set. The last two publications<sup>27,54</sup> mainly contain CCS values for organophosphorus flame retardants, compounds with a phosphate structure, which are common additives used in plastic materials. Since only three organophosphorus flame retardants were included in the previous self-built CCS database,<sup>43</sup> the addition of the CCS values from these two publications significantly expanded the chemical diversity of the current study.

CCS values for some compounds appeared in more than one publication. In such cases, the CCS data were rationalized as follows:

- (1) Chemical information retrieval: information including the compound identifier (CID), monoisotopic mass, molecular formula, canonical SMILES, and InChIKey of each CCS record was retrieved from PubChem using the R package *webchem*.<sup>56</sup>
- (2) Calculation of median CCS values for duplicated records: in the cases where different names were used for the same compound in the different publications, the InChIKey was used as a unique identifier. The median and relative standard deviation (RSD) of multiple CCS values were calculated, and the median CCS values were used in the model.

A total of 1721 CCS values were retained after the consolidation of duplicate records, which included 1076 CCS values for [M + H]<sup>+</sup> ions and 645 CCS values for [M + Na]<sup>+</sup> ions. In consolidated data, the CCS values of 248 [M + H]<sup>+</sup> ions (23.0%) and 72 [M + Na]<sup>+</sup> ions (11.2%) were median values of multiple CCS records. The CCS data set rationalization was performed using the R package *tidyverse*,<sup>57</sup> and the chemical class of each compound contributing to the model was obtained from ClassyFire.<sup>58</sup>

**2.2. Calculation and Selection of Molecular Descriptors.** MDs play a crucial role in the prediction of CCS values. In this work, three types of MDs were calculated using OCHEM<sup>59</sup> and ChemDes.<sup>60</sup> More information about MDs is shown in Supporting Information.

Descriptors that have a constant value or very few unique values relative to the number of samples have variance values equal or close to zero. Such descriptors contain little information and were considered less important for the model and excluded from the data set. Correlation coefficients ( $r$ ) between individual MDs and CCS were subsequently calculated, and only the MDs for which  $r > 0.6$  were retained. The remaining descriptors were auto-scaled to normalize the effect of the magnitude. The alvaDesc MDs were further rationalized by considering Extreme Gradient Boosting (XGBoost) importance. In XGBoost, the contribution of each variable to the model is calculated with respect to the number of times the variable is selected for splitting, weighted by the squared improvement to the model as a result of each split. The variable importance is then averaged across all the decision trees within the model.<sup>61</sup> In this study, the alvaDesc MDs accounting for 99 and 95% of the total XGBoost importance were retained.

**2.3. Development of the CCS Prediction Model.** For both  $[M + H]^+$  and  $[M + Na]^+$  ions, the data were randomly divided into training and testing sets in the ratio of 7:3. The training set was used for the calibration and optimization of the model, and the testing set was used for the external validation. The comparison of CCS prediction accuracy between various models (models developed with different algorithms and MDs in this study as well as public CCS prediction tools) was based on the testing set data. The R code for model building was provided in GitHub (<https://github.com/songxuechao/plasticCCS>).

In addition to the CCS data and descriptors, the machine learning algorithm employed was another important factor, affecting the predictive performance of the model. In this study, two algorithms that are often used for CCS prediction were compared: XGBoost and SVM. XGBoost is an optimized distributed gradient boosting library designed to be highly efficient and flexible<sup>62</sup> and was used to develop CCSSondemand.<sup>45</sup> The XGBoost model tuning consisted of 576 combinations of five important model parameters: eta (0.01, 0.05, 0.1, 0.3), max\_depth (3, 5, 7), min\_child\_weight (1, 3, 5), subsample (0.6, 0.7, 0.8, 0.9), and colsample\_bytree (0.6, 0.7, 0.8, 0.9). All combinations were evaluated using the training data set by a 10-fold cross validation. The optimal value of the *nrounds* parameter which controls the maximum number of iterations was returned using the minimized root-mean-square error of cross validation (RMSECV). Finally, the XGBoost model was built using the training data set with the optimized combination of parameters using the R package *xgboost*. The importance of MDs in the model was also calculated.

SVM is also a commonly used machine learning algorithm and has previously been used for the prediction of CCS values.<sup>44,63</sup> In this study, SVM with the radial basis function kernel was used to build the model. Two important hyperparameters were optimized in order to get accurate predictions: cost of constraints violation ( $C$ ) and gamma ( $\gamma$ ). The  $C$  parameter trades off the predictive performance of the training set against the model's margin, while the  $\gamma$  parameter defines how far the influence of a single training example reaches. Eight groups of  $C$  values (0.001, 0.005, 0.01, 0.025, 0.05, 0.1, 0.25, 0.5)/ $N_{MD}$  (i.e. number of MDs) and nine  $\gamma$  values ( $2^0$  to  $2^8$ ) formed 72 parameter combinations, which were then evaluated using 10-fold cross validation on the training set. The parameter combination providing the

minimum RMSECV was used in the SVM model using the R package *e1071*.

The performance of the models was assessed by comparing the following parameters: the coefficient of determination of the prediction ( $R_p^2$ ), the root-mean-square error of the prediction (RMSEP), the median relative error (MRE), and the percentage of molecules with relative deviations from experimental CCS values of less than 2, 3, and 5%.

The prediction performance of our model was compared to three publicly available CCS prediction tools: CCSSondemand (<https://ccs.on-demand.waters.com>) from Broeckling and co-workers,<sup>45</sup> AllCCS (<http://allccs.zhulab.cn>) from Zhu lab,<sup>35</sup> and CCSbase from Xu lab (<http://ccsbase.net>).<sup>46</sup>

**2.4. Prediction of CCS Values for Compounds in CPPdb and FCCdb.** The CPPdb consists of 4283 substances associated with plastic food packaging. The data set was rationalized by removing the metals and salts together with any substances with same InChIKey (replicates). Finally, only substances with a neutral mass between 50 and 1200 were retained. After following this procedure, 2883 substances from the CPPdb were retained. The FCCdb data set was also rationalized using the procedure described above, leading to 6508 substances retained in data set. The CCS values of the compounds retained from the databases were then predicted using the model that yielded the best performance in this study. Meanwhile, the chemical space covered by CPPdb, FCCdb, and our collected molecules was compared.

**2.5. Application of Predicted CCS Values to the Analysis of Plastic-Related Chemicals in Ebro River Water.** 2L of surface water were sampled from the Ebro River near the urban areas of Zaragoza, Spain. The river water was stored in an amber glass bottle and treated on the day of collection, using the previously developed procedures.<sup>30</sup> The final samples were analyzed using a Vion IMS-QToF mass spectrometer. The detailed procedures of sample treatment and operating conditions of the Vion are given in the [Supporting Information](#). The features ( $m/z$ \_RT\_CCS pairs), obtained from Vion IMS-QToF, were then screened against two plastic-related databases, CPPdb (2883 compounds) and FCCdb (6508 compounds), containing  $m/z$  values, adducts, and predicted CCS values. The  $m/z$  deviations of the measured values were less than 5 ppm as for CCS deviation, the filter setting was based on its prediction accuracy.

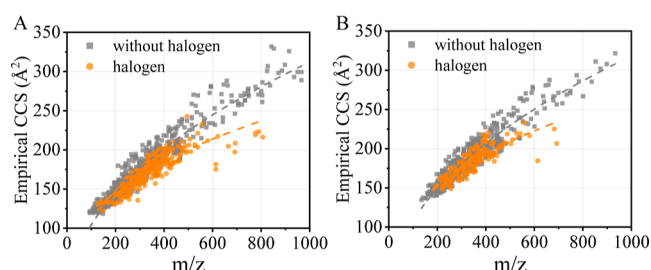
### 3. RESULTS

**3.1. CCS Data Set.** A total of 1076 and 645 CCS values were collated for  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively. CCS values ranged from 118.6 to 332.2 Å<sup>2</sup> for the  $[M + H]^+$  data and from 134.7 to 321.9 Å<sup>2</sup> for the  $[M + Na]^+$  data. Using ClassyFire,<sup>58</sup> the compounds were categorized into 10 super classes for the  $[M + H]^+$  adduct and 11 super classes for the  $[M + Na]^+$  adduct. The principal super classes assigned were benzenoids, organoheterocyclic compounds, lipids and lipid-like molecules, and organic acids and derivatives ([Figure S1](#)). Benzenoids include compounds commonly detected in plastics such as phthalate-based plasticizers, antioxidants, bisphenols, primary aromatic amines, and pesticides.

248 and 72 duplicate CCS values were found for  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively, across the seven publications, and the RSDs of the measurements are shown in [Figure S2](#). The RSD variation is less than 2% for 89.1% (221/248) of the  $[M + H]^+$  adducts of the molecules and 95.8% (69/72) of the  $[M + Na]^+$  adducts. Consequently, there

are 27 and 3 CCS values with RSDs higher than 2% for the  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively, and the measurements contributing to these values are summarized in Tables S2 and S3. The majority of CCS values with RSDs greater than 2% were obtained from the publications of Bijlsma et al. (2017),<sup>38</sup> Celma et al. (2020),<sup>29</sup> and Regueiro et al. (2016).<sup>39</sup> It appears that pesticide and drug-like compounds are more likely to produce a high variation of CCS values. Such compounds include picroxystrobin, acetopromazine, prochloraz, and oxadixyl, with the variation of the CCS measurements for the last two compounds being more than 20 Å<sup>2</sup>. The limit of CCS reproducibility, presence of protomers, and inconsistent CCS calibration across different instrument systems are three possible sources of deviations in CCS measurements. A more detailed explanation is given in the Supporting Information.

CCS is a value related to the size, shape, and charge of a molecule and understandably, CCS is also strongly correlated with the  $m/z$  value of a compound.<sup>27,31,41,54</sup> The correlation between  $m/z$  and the CCS value of the compounds considered in this study is shown in Figure 1. In general, the relationship



**Figure 1.** Empirical CCS vs  $m/z$  for (A)  $[M + H]^+$  and (B)  $[M + Na]^+$ .

between  $m/z$  and CCS can be described by a power regression model. The inclusion of more halogenated compounds in this study (a total of 302 and 149 halogenated molecules were included for  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively), highlighted a distinct difference in their  $m/z$  and CCS relationship when compared to the relationship for non-halogenated compounds. The halogenated compounds tended to have smaller CCS values for a given  $m/z$ . It is believed that halogens have a lower atomic radius per atomic mass unit in comparison to other elements, such as C, H, O, and N. The partially orthogonal structural information provided by CCS is discussed in the Supporting Information.

Some CCS values collated in this study were measured using drift tube IMS (DTIMS),<sup>27</sup> and deviations between <sup>TM</sup>CCS<sub>N<sub>2</sub></sub> and <sup>DT</sup>CCS<sub>N<sub>2</sub></sub> have previously been observed.<sup>53</sup> Since accurate CCS values are fundamental to obtain a reliable CCS prediction model, the <sup>DT</sup>CCS<sub>N<sub>2</sub></sub> values were compared to <sup>TM</sup>CCS<sub>N<sub>2</sub></sub> values available in the literature (Tables S4 and S5). 16 <sup>TM</sup>CCS<sub>N<sub>2</sub></sub> values were found in the literature that could be directly compared to <sup>DT</sup>CCS<sub>N<sub>2</sub></sub> values, and most of these values were for compounds in the types of plasticizers and organophosphorus flame retardants. Table S4 shows that 81.3% of the values agree to within 2% and the deviations ranged from 0.11% (for atrazine) to 2.88% (for tri-*n*-butyl phosphate) with an average of 1.15%. In the case of the  $[M + Na]^+$  adduct, 75.0% of the values agree to within 2%, and the deviations ranged from 0.15% (for di-*n*-butyl phosphate) to 4.23% [for mono(2-ethylhexyl) adipate], with an average of 1.32%. The median of the <sup>TM</sup>CCS<sub>N<sub>2</sub></sub> and <sup>DT</sup>CCS<sub>N<sub>2</sub></sub> values was

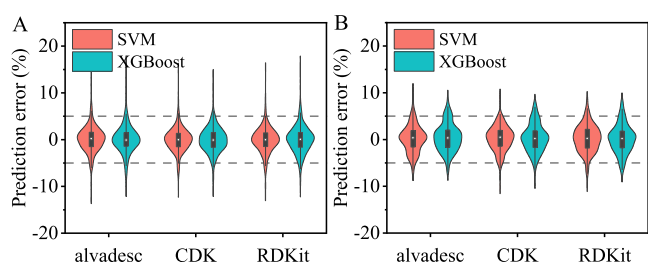
used when building the model to reduce any outlier measurements arising from the use of different IMS technologies.

**3.2. Selection and Weighting of Molecular Descriptors.** The selection of MDs can reduce training time, simplify the prediction model, and avoid overfitting; however, it is possible that meaningful information can also be lost, leading to a decrease in accuracy. For this reason, it is necessary to achieve a balance between the simplicity and accuracy of the model.

The numbers of MDs retained after each step of variable selection are shown in Figure S3, and the comparison of the model performance before and after variable selection is presented in Figure S4 and Tables S6–8. For alvaDesc MDs, the first 316 and 72 descriptors accounted for 99 and 95% of the total importance for  $[M + H]^+$  adducts. When the number of MDs was decreased from 1528 to 72, both the SVM and XGBoost models showed a slight decrease in the performance. The  $R_p^2$  of the SVM model decreased from 0.9802 to 0.9737, RMSEP increased from 4.47 to 5.43, and MRE increased from 1.50 to 1.52%. Considering that the model was significantly simplified and the performance was still acceptable, the 72 most important alvaDesc MDs were selected for the  $[M + H]^+$  adduct data. In the case of the  $[M + Na]^+$  adduct CCS predictions, the models based on the first 193 MDs showed a comparable performance with the models built on 1361 MDs. Therefore, the 193 most significant MDs were selected for  $[M + Na]^+$  adduct data.

On determining the descriptors using CDK and RDKit, after the elimination of MDs that show low correlation with CCS ( $r < 0.6$ ), 84 and 65 CDK descriptors and 33 and 27 RDKit descriptors were retained for  $[M + H]^+$  adducts and  $[M + Na]^+$  adducts, respectively; they were not filtered further. Table S7 shows that 84 CDK MDs can provide accurate prediction results for  $[M + H]^+$  adducts. A remarkable reduction in the performance of the model was observed for  $[M + Na]^+$  adducts, when the number of MDs was reduced from 207 to 65. Therefore, 84 and 207 CDK MDs were selected for the  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively. In the case of RDKit, 33 and 125 MDs were retained for  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively, based on the performance of the model (Table S8).

**3.3. Model Performance.** After dividing the collated CCS values into a training data set and a testing data set, 329 and 181 CCS values were included in the testing set for  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively. For each adduct, six CCS prediction models were developed based on the combinations of two algorithms (XGBoost and SVM) and three types of MDs (alvaDesc, CDK, and RDKit). The distribution of prediction errors and model parameters for each model are shown in Figure 2 and Table 1, respectively. In the case of the  $[M + H]^+$  adducts, more than 90% of molecules showed prediction errors within 5% for all six models. The SVM-based model in conjunction with the CDK descriptors provided the best predictive performance.  $R_p^2$  and MRE were 0.9786 and 1.42%, respectively, and more than 93 and 64% of molecules had prediction errors of less than 5 and 2%, respectively. This model also provided a better predictive performance for the  $[M + Na]^+$  adduct with more than 95 and 58% molecules having prediction errors of less than 5 and 2%, respectively. The results also show that the model for  $[M + H]^+$  adducts should use a different set of descriptors to those used for the model for  $[M + Na]^+$  adducts, implying that a unique CCS



**Figure 2.** Violin-plot illustrating the prediction errors of the SVM and XGBoost models using different sets of descriptors: (A)  $[M + H]^+$  and (B)  $[M + Na]^+$ .

prediction model should be developed for each adduct. This is highlighted by the studies of Bijlsma et al. (2017),<sup>38</sup> in which a single set of descriptors was used for the CCS prediction of all positive ions and demonstrated that the CCS values predicted for  $[M + Na]^+$  adducts were less accurate in general.

In comparison to our previous study,<sup>43</sup> a more accurate prediction of CCS values for  $[M + Na]^+$  adducts is achieved here. The value of RMSEP decreased from 8.2 to 5.5 Å<sup>2</sup>, the percentage of molecules with prediction errors less than 5% increased from 81.3 to 95.0% and those with prediction errors less than 2% increased from 54.7 to 58%. Even though there is a dramatic improvement, the prediction of CCS values for  $[M + Na]^+$  adducts was still less accurate than that for  $[M + H]^+$  adducts. It is believed that the main reason for this is that the MDs are calculated from neutral molecules, and a sodium adduct can lead to a more diverse range of molecular conformations in 3D space compared to protonation.<sup>38,41,64</sup> One way to improve the accuracy of CCS predictions would be to determine the descriptors for ionized molecules, rather than neutral molecules. However, such an approach is more complicated and computationally expensive, in addition, conformational analysis is always required before the calculation of the descriptors.<sup>65</sup>

The CCS prediction for halogenated molecules was also more accurate using the current SVM model compared to our previous study.<sup>43</sup> 95.3% (81 out of 85 molecules) and 65.9% (56 out of 85 molecules) of protonated halogenated molecules had prediction errors of less than 5 and 2%, respectively. This compares to our previous study<sup>43</sup> for which the percentages were only 86.7 and 40%, respectively. This significant improvement could be due to the additional halogenated molecules in the training set, which supports previous observations that structure similarity between predictions and

the training set significantly affect the accuracy of CCS predictions.<sup>35</sup> To further validate this conjecture, we excluded the 217 halogenated molecules from the training set for  $[M + H]^+$  adducts, leaving 530 non-halogenated molecules to rebuild the SVM model for the prediction of CCS values for molecules in the testing set. A comparison of CCS prediction results, with and without halogenated compounds in the training set, is shown in Figure S5. It is evident that upon excluding halogenated compounds from the training set, the prediction errors for the 244 non-halogenated compounds in the test data are similar to those generated when the halogens were included in the training data. However, the predicted CCS values of 85 halogenated compounds in the test data has significantly larger errors when the halogens were excluded from the training data: MRE increased from 1.46 to 1.87%, and the proportion of halogenated compounds with prediction errors <2% decreased from 65.9 to 54.1%. This confirms that the chemical diversity of training set is an important factor, which affects the prediction accuracy for the test data.

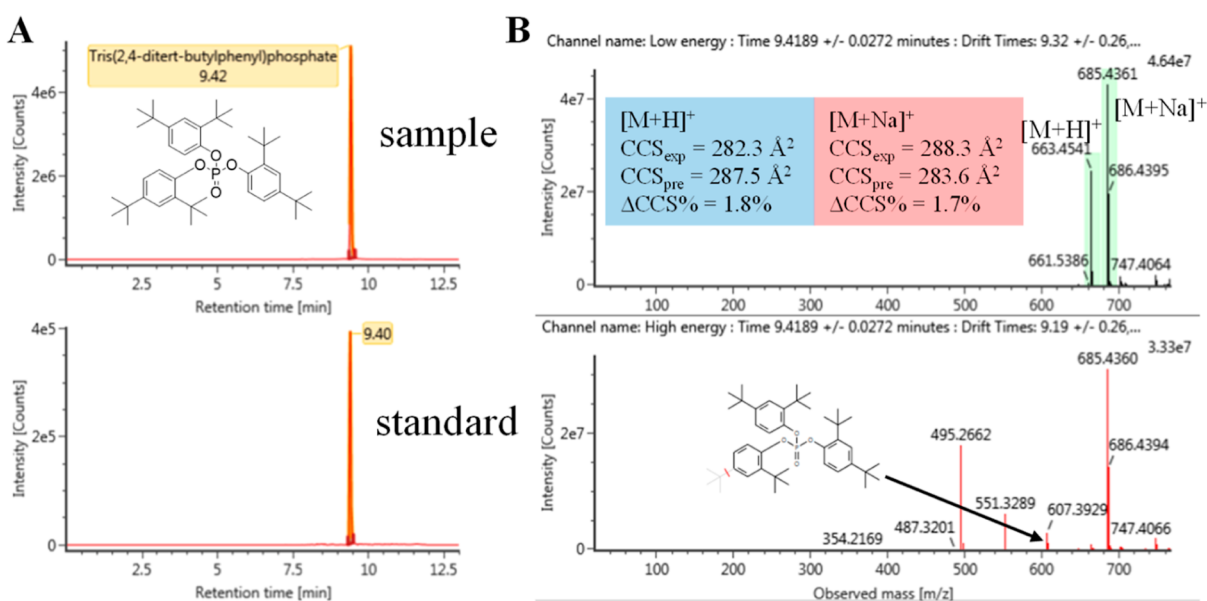
The protonated molecules for which the prediction error in the CCS value was greater than 5% were further investigated. The presence of protomers can lead to high CCS prediction errors. For example, two different CCS values (160.5 and 176.2 Å<sup>2</sup>) have been reported for acetopromazine in previous studies,<sup>29,38</sup> the predicted CCS value of 179.8 Å<sup>2</sup> matched well with the CCS value of the more extended protomer. Similar behavior was also observed in the work of Zhou et al.<sup>44</sup> More discussions are given in the Supporting Information.

Through the comparison of the six models and the comparison with our previous study,<sup>43</sup> the SVM model based on CDK MDs provided the most accurate predictions. The chemical diversity of the training set seems to be a more crucial factor for CCS prediction than descriptors and algorithms. The possibility of multiple protomers is another important factor, affecting the accuracy since only one predicted CCS value can currently be determined for a given adduct by machine learning models. Besides, we opted to use SVM due to its easy configuration with few hyperparameters, as well as its ability to provide reproducible prediction results.

**3.4. Comparison between the SVM Model and Public CCS Prediction Tools.** The outcomes from the SVM model based on CDK MDs were compared to those from three publicly available CCS prediction tools: CCSondemand, AllCCS, and CCSbase. The distributions of the prediction errors for all models are illustrated in Figure S6, and the

**Table 1.** Performance of the Models Developed Using Different Descriptors and Algorithms

adducts	descriptor	algorithm	$R_p^2$	RMSEP	<2%	<3%	<5%	MRE (%)
$[M + H]^+$	alvaDesc	SVM	0.9737	5.43	61.7	79.0	91.8	1.52
		XGBoost	0.9727	5.53	61.7	75.7	90.6	1.44
	CDK	SVM	<b>0.9786</b>	<b>4.90</b>	<b>64.7</b>	<b>82.7</b>	<b>93.3</b>	<b>1.42</b>
		XGBoost	0.9765	5.14	59.6	78.7	94.2	1.61
	RDKit	SVM	0.9772	5.09	63.8	79.6	93.0	1.46
		XGBoost	0.9700	5.80	58.1	74.2	90.3	1.58
$[M + Na]^+$	alvaDesc	SVM	0.9570	5.83	54.1	67.4	90.1	1.81
		XGBoost	0.9593	5.76	52.5	72.9	89.0	1.88
	CDK	SVM	<b>0.9618</b>	<b>5.53</b>	<b>58.0</b>	<b>74.6</b>	<b>95.0</b>	<b>1.76</b>
		XGBoost	0.9555	5.95	53.0	68.5	90.1	1.81
	RDKit	SVM	0.9511	6.18	49.2	72.9	90.1	2.01
		XGBoost	0.9577	5.82	53.6	69.1	87.8	1.81



**Figure 3.** Identification of tris(2,4-ditert-butylphenyl)phosphate. (A) Extracted ion chromatograms from the sample and standard, (B) low- and high-energy spectra, fragment assignment, comparison between experimental and predicted CCS values.

corresponding MRE for each chemical class is shown in Figure S7.

The CCS values of 65 and 74% of protonated molecules were predicted with an error of less than 2% by SVM and CCSondemand, respectively. More than 93% of protonated molecules has prediction errors less than 5% for both models. CCSondemand was trained by approximately 7325 experimental <sup>TW</sup>CCS<sub>N<sub>2</sub></sub> values obtained from 3775 compounds.<sup>45</sup> The training data set contains CCS values of chemicals found in plastic food packaging and pesticides, so when the CCS values of such molecules are predicted by CCSondemand, one would expect smaller prediction errors. The predictive capabilities of AllCCS and CCSbase were not as good as those for SVM and CCSondemand for the compounds considered in this study. This is possibly due to the dissimilarity of the structures of chemicals in plastics and the molecules used in the training sets of AllCCS and CCSbase.

The results for [M + Na]<sup>+</sup> adducts showed that the SVM model gave more accurate predictions than the other tools. The enhanced performance of SVM is possibly due to the higher number of MDs used in this model ( $n = 207$ ) as only 15 MDs were used in AllCCS.<sup>35</sup> More detailed comparison is shown in the Supporting Information.

The AllCCS tool is also based on the SVM algorithm and CDK MDs; however, there are two main differences between AllCCS and our model: the training data and the number of MDs. In order to investigate which factor leads to the significantly different prediction results between AllCCS and our model, we built a SVM model based on our CCS training data and the 15 MDs used for AllCCS and compared their prediction results for the testing set to those obtained from our original model and AllCCS (Table S9). For both [M + H]<sup>+</sup> and [M + Na]<sup>+</sup> adducts, less accurate prediction results were obtained from SVM models based on 15 MDs than with our original SVM model. MRE values increased from 1.4 to 1.6%, for [M + H]<sup>+</sup> adducts and 1.8 to 2.1% for [M + Na]<sup>+</sup> adducts. The results from the SVM model based on 15 MDs and the AllCCS tool using the same MDs but different training data can be seen in Table S9. AllCCS shows significantly larger

prediction errors, with MRE values of 2.2% for [M + H]<sup>+</sup> adducts and 3.3% for [M + Na]<sup>+</sup> adducts. These results show that the data used to train the model have a greater effect on the prediction accuracy of the model than the MDs.

These results show that, when compared to other available prediction tools, the SVM model based on the CDK MDs can improve the prediction of CCS values, especially for sodiated molecules. The CCS values for the [M + H]<sup>+</sup> and [M + Na]<sup>+</sup> adducts of the molecules in CPPdb and FCCdb were subsequently predicted by the SVM model developed here. The two databases were then transformed into screening libraries, which were used for the suspect screening of plastic-related chemicals in Ebro River water.

**3.5. Plastic-Related Chemicals Tentatively Identified in Ebro River Water.** Approximately 95% of predicted CCS values (93.3% for [M + H]<sup>+</sup> adducts and 95.0% for [M + Na]<sup>+</sup> adducts) are within 5% deviation with respect to experimental values. Thus, the tolerance for CCS deviations was set as 5% in the suspect screening of plastic-related chemicals in Ebro River water. Two main aspects of using predicted CCS values in the identification of unknowns were investigated: reducing the number of false positives and increasing the confidence level of identified compounds. The river water samples were screened against 9391 compounds in CPPdb and FCCdb to search for plastic-related chemicals. The number of candidates with and without the confirmation of CCS values was compared. The addition of the CCS filter decreased the number of candidate compounds from 376 to 204 (45.7%).

A total of 98 plastic-related chemicals were tentatively identified in the Ebro River surface water samples from the CPPdb and FCCdb databases, of which 26 compounds were confirmed using reference standards. The tentatively identified compounds consisted of 12 plasticizers, 10 flame retardants, 6 antioxidants, 9 slip agents, 10 dyes, and 26 surfactants (including glycol and glycerol derivatives). NIAS were also detected in Ebro River water, including the ethylene terephthalate cyclic trimer, a common oligomer of polyethylene terephthalate,<sup>66</sup> and bisphenol A bis(2,3-dihydroxypropyl) ether, a hydrolysis product of bisphenol A diglycidyl

ether.<sup>67</sup> Detailed information about the identified compounds is available in the [Supporting Information](#).

The most abundant compound detected in the Ebro River water samples was tris(2,4-ditert-butylphenyl)phosphate. This is a degradation product of Irgafos 168 (a commonly used phosphite antioxidant in plastics).<sup>68</sup> Previous studies have shown that tris(2,4-ditert-butylphenyl)phosphate was an abundant contaminant in indoor dust<sup>21</sup> and fine particulate matter.<sup>14</sup> The predicted CCS values for tris(2,4-ditert-butylphenyl)phosphate had deviations less than 2% versus the experimental values ([Figure 3](#)).

The benefit of predicted CCS values in identification of unknowns is more relevant either when the analyte is at low concentration levels or the reference standard is not available. 1,4,7-Trioxacyclotridecane-8,13-dione is a reaction product from adipate plasticizer/adipate acid and ethylene glycol, and its molecular structure and mass spectra are shown in [Figure S8](#). The figure shows the isotopic pattern for the  $[M + H]^+$  adduct was indistinct, and no fragment ions were observed in the high-energy spectrum for the compound, possibly due to the low concentration and ineffectual assignment of fragment ions to the respective precursor ions. A fragment ion at  $m/z$  value 155.0699 was observed in low-energy spectrum, which corresponds to the loss of the ethylene glycol unit. CCS deviation for the  $[M + H]^+$  and  $[M + Na]^+$  adducts of 1,4,7-trioxacyclotridecane-8,13-dione was 0.4 and  $-2.1\%$ , respectively. In this case, even though no abundant fragmentation information was obtained for 1,4,7-trioxacyclotridecane-8,13-dione, the combination of RT,  $m/z$  and CCS contribute to a reliable identification.

In some cases, even when the analyte is at a high concentration in the sample, fragments ions may still not be assigned in the high-energy mass spectrum, as a result of rigid structures of less labile molecules. An example of this is given in [Figure S9](#), in which the mass spectra of Antiblaze V6, a flame retardant in plastics, are shown. The fragment ions observed in the high-energy spectrum are in low abundance, and substructures of the parent molecule could not be assigned. The predicted CCS value ( $207.9 \text{ \AA}^2$ ) had a 3.5% deviation when compared to the experimental value ( $215.2 \text{ \AA}^2$ ). Additionally, an experimental CCS value ( $211.4 \text{ \AA}^2$ ) for Antiblaze V6 was found in the literature<sup>27</sup> and has a deviation of  $-1.8\%$  from our experimental value. It is not possible to confirm the identification of this compound due to the lack of a reference standard; however, the comparison between predicted and experimental CCS values, the  $m/z$  values, and the characteristic chlorine isotopic pattern provides high confidence for the assignment.

False positive assignments have been observed for which the tolerances of  $m/z$  error  $<5$  ppm and CCS deviation  $<5\%$  are satisfied. For example, the ion with  $m/z$  327.0785 and CCS  $169.9 \text{ \AA}^2$  at a RT of 9.42 min is a good match for triphenyl phosphate. However, the reference standard was detected with a RT of 6.50 min, thereby showing this assignment to be a false positive. The addition of RT predictions may be able to eliminate this kind of false positives, as shown by previous studies.<sup>67,68</sup>

## 4. DISCUSSION

**4.1. Suitability of Combining Both  $^{DT}CCS_{N_2}$  and  $^{TW}CCS_{N_2}$  Values in the Model.** CCS values measured using DTIMS can differ from those measured using traveling wave IMS (TWIMS) platforms.<sup>53</sup> Therefore, the suitability of

combining both  $^{DT}CCS_{N_2}$  and  $^{TW}CCS_{N_2}$  values in the CCS prediction model was investigated. 16 compounds have both  $^{DT}CCS_{N_2}$  and  $^{TW}CCS_{N_2}$  values for both  $[M + H]^+$  and  $[M + Na]^+$  adducts ([Tables S4 and S5](#)). Measurements of  $^{DT}CCS_{N_2}$  alone are present for 39  $[M + H]^+$  adducts and 65  $[M + Na]^+$  adducts. Of these, 27  $^{DT}CCS_{N_2}$  values for  $[M + H]^+$  adducts and 51  $^{DT}CCS_{N_2}$  values for  $[M + Na]^+$  adducts are present in the training data. The  $^{DT}CCS_{N_2}$  values were removed from the training data, the SVM models were rebuilt, and their performance was compared to the original SVM models ([Table S10](#)). The prediction accuracy of the CCS values for the  $[M + H]^+$  adducts remained similar to the original results; however, the predicted CCS values for the  $[M + Na]^+$  adducts were less accurate. RMSEP increased from 5.5 to  $5.8 \text{ \AA}^2$ , and the proportion of compounds with prediction errors was  $<2\%$  decreased from 58.0 to 55.8%. The reduction in the prediction accuracy upon removing the  $^{DT}CCS_{N_2}$  values from the training data is probably due to the reduction in the diversity of chemical structures. The  $^{DT}CCS_{N_2}$  values were mainly for organophosphate flame retardants and phthalate monoesters, both of which are additives commonly used in plastics.<sup>11,13</sup>

It should be noted that in this study, the differences between  $^{DT}CCS_{N_2}$  and  $^{TW}CCS_{N_2}$  values are relatively small. Higher CCS deviations were observed between  $^{TW}CCS_{N_2}$  values from different laboratories than between  $^{DT}CCS_{N_2}$  and  $^{TW}CCS_{N_2}$  values ([Tables S2–S5](#)). Based on these observations, we decided to use both  $^{DT}CCS_{N_2}$  and  $^{TW}CCS_{N_2}$  values in the training data of the model.

**4.2. Weighting and Collinearity of CDK MDs.** The important CDK MDs for the prediction of the CCS values are shown in [Figures S10 and S11](#), and a brief description of these important MDs is also given in [Table S11](#) and Supplemental Results and Discussion. The effect of collinearity between CDK MDs was investigated by building models that omitted highly correlated MDs. The variance inflation factor (VIF) is a measure of the correlation between MDs with higher values indicating greater correlation. Two models were built to study collinearity of MDs, one for which MDs with a VIF value greater than 50 were excluded and one for which MDs with a VIF value greater than 20 were excluded. A comparison of the predictive performance between the two new models and the original model is shown in [Figure S12](#). The figure shows that for  $[M + H]^+$  adducts, 33 MDs with a VIF value below 50 were retained and 24 MDs with VIF value below 20 were retained. The reduction in the number of the MDs slightly decreased the prediction accuracy of both the SVM and XGBoost models. Similar behavior was observed when the same procedure was applied to the models for the  $[M + Na]^+$  adduct. Since the models built with 84 and 207 CDK MDs for the  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively, provide more accurate predictions, and the complexity of the model was still deemed to be acceptable, the number of MDs was not reduced in the final models.

**4.3. Approaches to Improve the Prediction Accuracy.** There are several ways in which the accuracy of CCS predictions could potentially be improved. First, more experimental CCS values can be collected for the training set to increase the chemical diversity and universality of the model. A total of 17 and 15 chemical super classes are considered in CCSONdemand and AllCCS respectively, while in this study, only 10 and 11 super classes are covered by 1076 and 645 CCS values for  $[M + H]^+$  and  $[M + Na]^+$  adducts, respectively. [Table S12](#) presents the 50 compounds in CPPdb

and FCCdb that were not covered by the chemical space of our collected CCS records. Generally, these compounds have high molecular mass and contain long linear-chain structures.

Second, MDs based on ionized molecules could improve predictions, especially for  $[M + Na]^+$  adducts. There is a much bigger difference between the structural conformation of sodiated and neutral molecules than there is between protonated and neutral molecules.<sup>64</sup> This makes it difficult to obtain accurate predicted CCS values for sodiated molecules when the descriptors are derived from neutral molecules. Taking into account that deriving MDs from ionized molecules is time-consuming and complex, and as such, MDs derived from neutral molecules are probably sufficiently accurate for  $[M + H]^+$  adducts. In the study by Gonzales et al. (2016),<sup>48</sup> MDs of deprotonated phenolics were determined for a CCS prediction model, and 92.8% (52/56) of molecules was predicted within 5% of their measured values. In the present study, a similar proportion (93.3%) of protonated molecules was predicted with an error less than 5%, highlighting that MDs determined from neutral molecules are sufficient for the accurate prediction of CCS values of protonated molecules.

Third, improving the reproducibility of commercially available IMS devices such as TWIMS and DTIMS will lead to more precise and accurate CCS measurements, which, when used as inputs to prediction models, will improve the performance of the models. At the time of writing, commercial IMS devices have relatively low reproducibility, which makes it impractical to adopt an accuracy threshold lower than 2% when matching measured CCS values to library values within a suspect screening workflow.<sup>41,42,69</sup>

**4.4. Current Limitations and Future Prospects.** In this study, the CCS prediction models were only built for positive ions. This is understandable to some extent as most additives in plastic products, such as plasticizers, antioxidants, flame retardants, photoinitiators, and slip agents, are detected in the positive ion mode.<sup>20</sup> In some cases, compounds can be only detected, or show a higher response, in the negative ion mode. Such compounds include lubricants (lauric acid and oleic acid) and surfactants (perfluorooctanesulfonic acid and perfluorobutanesulfonic acid), which were detected in the Ebro River water samples using our in-house plastic additives library (see the [Supporting Information](#)). Therefore, a CCS prediction model for negative ions needs to be developed in the future. Additionally, the CCS prediction models developed herein are only available to a small set of privileged users, and work needs to be undertaken to develop them into an open-access tool.

Many emerging contaminants associated with plastics, such as tricaprins, polyethylene glycol, and polypropylene glycol oligomers, do not exist in the CPPdb and FCCdb databases. These compounds were detected at high abundance in the Ebro River water samples using our in-house plastic additives library. With the rapid growth of newly reported plastic-related chemicals, the CPPdb and FCCdb databases need to be continuously expanded and updated. The construction of an integrated plastic-related database containing name, adducts,  $m/z$  values, predicted CCS values, and predicted RTs will facilitate the identification of extractables and leachables from plastics in HRMS-based screening strategies.

In summary, the SVM model, based on CDK descriptors presented here, provided more accurate CCS predictions than the XGBoost algorithm and other descriptors. The CCS values of 93.3%  $[M + H]^+$  adducts and 95.0%  $[M + Na]^+$  adducts were predicted within 5% of their measured values. It has been

shown that the chemical diversity of the training set appears to have more influence on the predictive performance than alternative algorithms and MDs investigated here. Indeed, CCS predictions for halogenated compounds were more accurate following the incorporation of more CCS records of halogenated compounds into the training set. Increasing the number of experimental CCS values and improving the reproducibility of CCS measurements seem to be two feasible ways to further increase the performance of prediction models. In future work, a CCS prediction model for negative ions will be developed, and work toward making all models open-access will be undertaken.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.est.2c02853>.

Chemical classes of collected molecules; RSDs of CCS values; variable selection processes; optimization of MDs; effect of halogenated compounds; comparison between SVM models to other public tools; identification of 1,4,7-trioxacyclotridecane-8,13-dione; mass spectra of Antiblaze V6; influential CDK descriptors; correlation between CCS and Atomic and Bond Contributions of van der Waals volume (VABC); effect of collinearity; collected experimental CCS values; CCS records with RSD higher than 2%; comparison between  $^{DT}CCS_{N_2}$  and  $^{TW}CCS_{N_2}$ ; optimization of alvaDesc, CDK, and RDKit descriptors; comparison with AllCCS; comparison of SVM models before and after excluding  $^{DT}CCS_{N_2}$  values; calculation of MDs; river water treatment; conditions of Vion IMS-QTOF; sources of CCS deviations; reasons leading to high prediction errors; comparison between SVM to public tools; and important CDK descriptors ([PDF](#))

Collected empirical CCS values ([XLSX](#))

Plastic-related chemicals tentatively identified in Ebro River water ([XLSX](#))

## ■ AUTHOR INFORMATION

### Corresponding Author

Cristina Nerin – Department of Analytical Chemistry, Aragon Institute of Engineering Research I3A, CPS-University of Zaragoza, 50018 Zaragoza, Spain; [orcid.org/0000-0003-2685-5739](https://orcid.org/0000-0003-2685-5739); Phone: +34 976761873; Email: [cnerin@unizar.es](mailto:cnerin@unizar.es)

### Authors

Xue-Chao Song – Department of Analytical Chemistry, Aragon Institute of Engineering Research I3A, CPS-University of Zaragoza, 50018 Zaragoza, Spain

Nicola Dreolin – Waters Corporation, SK9 4AX Wilmslow, U.K.

Elena Canellas – Department of Analytical Chemistry, Aragon Institute of Engineering Research I3A, CPS-University of Zaragoza, 50018 Zaragoza, Spain

Jeff Goshawk – Waters Corporation, SK9 4AX Wilmslow, U.K.

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.est.2c02853>



## Author Contributions

X.-C.S.: conceptualization, methodology, software, investigation, model building, and writing—original draft. N.D.: software, equipment, and writing—review and editing. E.C.: supervision, conceptualization, and writing—review and editing. J.G.: software, equipment, library building, and writing—review and editing. C.N.: supervision, funding acquisition, and writing—review and editing. All authors have given approval to the final version of the manuscript.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

X.-C.S. acknowledges the grant received from the China Scholarship Council (201806780031). The authors thank Waters for access to an IMS-QTOF instrument and Gobierno de Aragón and Fondo Social Europeo for the financial help given to the GUIA group T53\_20R. The authors thank Spanish Ministry of Research and Innovation for the project RTI2018-097805-B-I00.

## ABBREVIATIONS

IMS	ion mobility separation
HRMS	high-resolution mass spectrometry
CCS	collision cross-section
SVM	support vector machine
MRE	median relative errors
CPPdb	Chemicals associated with Plastic Packaging Database
FCCdb	Food Contact Chemicals Database
MW	molecular weight
NIAS	non-intentionally added substances
PFAS	perfluoroalkyl substances
FCMs	food contact materials
<sup>TW</sup> CCS <sub>N2</sub>	traveling wave CCS
<sup>DT</sup> CCS <sub>N2</sub>	drift tube CCS
CID	compound identifier
RSDs	relative standard deviations
MD	molecular descriptor
<i>r</i>	correlation coefficients
XGBoost	extreme gradient boosting
RMSECV	root-mean-square error of cross validation
RBF	radial basis function
<i>C</i>	cost of constraint violation
$\gamma$	gamma
$R_p^2$	determination coefficient of prediction
RMSEP	root-mean-square error of prediction
RT	retention time
DTIMS	drift tube ion mobility separation
PET	polyethylene terephthalate
TWIMS	traveling wave ion mobility separation
VIF	variance inflation factor
PEG	polyethylene glycol
PPG	polypropylene glycol

## REFERENCES

- (1) Ilyas, M.; Ahmad, W.; Khan, H.; Yousaf, S.; Khan, K.; Nazir, S. Plastic waste as a significant threat to environment - a systematic literature review. *Rev. Environ. Health* **2018**, *33*, 383–406.
- (2) Geyer, R.; Jambeck, J. R.; Law, K. L. Production, use, and fate of all plastics ever made. *Sci. Adv.* **2017**, *3*, No. e1700782.

(3) Prata, J. C.; da Costa, J. P.; Lopes, I.; Duarte, A. C.; Rocha-Santos, T. Environmental exposure to microplastics: An overview on possible human health effects. *Sci. Total Environ.* **2020**, *702*, 134455.

(4) Hernandez, L. M.; Xu, E. G.; Larsson, H. C. E.; Tahara, R.; Maisuria, V. B.; Tufenkji, N. Plastic Teabags Release Billions of Microparticles and Nanoparticles into Tea. *Environ. Sci. Technol.* **2019**, *53*, 12300–12310.

(5) He, Y.-J.; Qin, Y.; Zhang, T.-L.; Zhu, Y.-Y.; Wang, Z.-J.; Zhou, Z.-S.; Xie, T.-Z.; Luo, X.-D. Migration of (non-) intentionally added substances and microplastics from microwavable plastic food containers. *J. Hazard. Mater.* **2021**, *417*, 126074.

(6) Biryol, D.; Nicolas, C. I.; Wambaugh, J.; Phillips, K.; Isaacs, K. High-throughput dietary exposure predictions for chemical migrants from food contact substances for use in chemical prioritization. *Environ. Int.* **2017**, *108*, 185–194.

(7) Hahladakis, J. N.; Velis, C. A.; Weber, R.; Iacovidou, E.; Purnell, P. An overview of chemical additives present in plastics: Migration, release, fate and environmental impact during their use, disposal and recycling. *J. Hazard. Mater.* **2018**, *344*, 179–199.

(8) Su, Q.-Z.; Vera, P.; Nerín, C.; Lin, Q.-B.; Zhong, H.-N. Safety concerns of recycling postconsumer polyolefins for food contact uses: Regarding (semi-)volatile migrants untargetedly screened. *Resour. Conserv. Recycl.* **2021**, *167*, 105365.

(9) Liu, R.; Mabury, S. A. Synthetic Phenolic Antioxidants: A Review of Environmental Occurrence, Fate, Human Exposure, and Toxicity. *Environ. Sci. Technol.* **2020**, *54*, 11706–11719.

(10) Liu, R.; Mabury, S. A. Identification of Photoinitiators, Including Novel Phosphine Oxides, and Their Transformation Products in Food Packaging Materials and Indoor Dust in Canada. *Environ. Sci. Technol.* **2019**, *53*, 4109–4118.

(11) Liu, X.; Peng, C.; Shi, Y.; Tan, H.; Tang, S.; Chen, D. Beyond Phthalate Diesters: Existence of Phthalate Monoesters in South China House Dust and Implications for Human Exposure. *Environ. Sci. Technol.* **2019**, *53*, 11675–11683.

(12) Liu, X.; Chen, D.; Yu, Y.; Zeng, X.; Li, L.; Xie, Q.; Yang, M.; Wu, Q.; Dong, G. Novel Organophosphate Esters in Airborne Particulate Matters: Occurrences, Precursors, and Selected Transformation Products. *Environ. Sci. Technol.* **2020**, *54*, 13771–13777.

(13) Liu, X.; Zeng, X.; Dong, G.; Venier, M.; Xie, Q.; Yang, M.; Wu, Q.; Zhao, F.; Chen, D. Plastic Additives in Ambient Fine Particulate Matter in the Pearl River Delta, China: High-Throughput Characterization and Health Implications. *Environ. Sci. Technol.* **2021**, *55*, 4474–4482.

(14) Shi, J.; Xu, C.; Xiang, L.; Chen, J.; Cai, Z. Tris(2,4-di-tert-butylphenyl)phosphate: An Unexpected Abundant Toxic Pollutant Found in PM<sub>2.5</sub>. *Environ. Sci. Technol.* **2020**, *54*, 10570–10576.

(15) González-Mariño, I.; Ares, L.; Montes, R.; Rodil, R.; Cela, R.; López-García, E.; Postigo, C.; López de Alda, M.; Pocurull, E.; Marcé, R. M.; Bijlsma, L.; Hernández, F.; Picó, Y.; Andreu, V.; Rico, A.; Valcárcel, Y.; Miró, M.; Etxebarria, N.; Quintana, J. B. Assessing population exposure to phthalate plasticizers in thirteen Spanish cities through the analysis of wastewater. *J. Hazard. Mater.* **2021**, *401*, 123272.

(16) Gong, X.; Zhang, W.; Zhang, S.; Wang, Y.; Zhang, X.; Lu, Y.; Sun, H.; Wang, L. Organophosphite Antioxidants in Mulch Films Are Important Sources of Organophosphate Pollutants in Farmlands. *Environ. Sci. Technol.* **2021**, *55*, 7398–7406.

(17) Bolívar-Subirats, G.; Cortina-Puig, M.; Lacorte, S. Multiresidue method for the determination of high production volume plastic additives in river waters. *Environ. Sci. Pollut. Res. Int.* **2020**, *27*, 41314–41325.

(18) Schmidt, N.; Fauvelle, V.; Ody, A.; Castro-Jiménez, J.; Jouanno, J.; Changeux, T.; Thibaut, T.; Sempéré, R. The Amazon River: A Major Source of Organic Plastic Additives to the Tropical North Atlantic? *Environ. Sci. Technol.* **2019**, *53*, 7513–7521.

(19) Bolívar-Subirats, G.; Rivetti, C.; Cortina-Puig, M.; Barata, C.; Lacorte, S. Occurrence, toxicity and risk assessment of plastic additives in Besos river, Spain. *Chemosphere* **2021**, *263*, 128022.

- (20) Nerin, C.; Alfaro, P.; Aznar, M.; Domeño, C. The challenge of identifying non-intentionally added substances from food packaging materials: a review. *Anal. Chim. Acta* **2013**, *775*, 14–24.
- (21) Liu, R.; Mabury, S. A. Unexpectedly High Concentrations of a Newly Identified Organophosphate Ester, Tris(2,4-di-tert-butylphenyl) Phosphate, in Indoor Dust from Canada. *Environ. Sci. Technol.* **2018**, *52*, 9677–9683.
- (22) Liu, R.; Mabury, S. A. Organophosphite Antioxidants in Indoor Dust Represent an Indirect Source of Organophosphate Esters. *Environ. Sci. Technol.* **2019**, *53*, 1805–1811.
- (23) Cecon, V. S.; Da Silva, P. F.; Curtzwiler, G. W.; Vorst, K. L. The challenges in recycling post-consumer polyolefins for food contact applications: A review. *Resour. Conserv. Recycl.* **2021**, *167*, 105422.
- (24) Curtzwiler, G. W.; Silva, P.; Hall, A.; Ivey, A.; Vorst, K. Significance of Perfluoroalkyl Substances (PFAS) in Food Packaging. *Integr. Environ. Assess. Manage.* **2021**, *17*, 7–12.
- (25) Dodds, J. N.; Alexander, N. L. M.; Kirkwood, K. I.; Foster, M. R.; Hopkins, Z. R.; Knappe, D. R. U.; Baker, E. S. From Pesticides to Per- and Polyfluoroalkyl Substances: An Evaluation of Recent Targeted and Untargeted Mass Spectrometry Methods for Xenobiotics. *Anal. Chem.* **2021**, *93*, 641–656.
- (26) Luo, Y.-S.; Aly, N. A.; McCord, J.; Strynar, M. J.; Chiu, W. A.; Dodds, J. N.; Baker, E. S.; Rusyn, I. Rapid Characterization of Emerging Per- and Polyfluoroalkyl Substances in Aqueous Film-Forming Foams Using Ion Mobility Spectrometry-Mass Spectrometry. *Environ. Sci. Technol.* **2020**, *54*, 15024–15034.
- (27) Belova, L.; Caballero-Casero, N.; van Nuijs, A. L. N.; Covaci, A. Ion Mobility-High-Resolution Mass Spectrometry (IM-HRMS) for the Analysis of Contaminants of Emerging Concern (CECs): Database Compilation and Application to Urine Samples. *Anal. Chem.* **2021**, *93*, 6428–6436.
- (28) Canellas, E.; Vera, P.; Nerin, C. Ion mobility quadrupole time-of-flight mass spectrometry for the identification of non-intentionally added substances in UV varnishes applied on food contact materials. A safety by design study. *Talanta* **2019**, *205*, 120103.
- (29) Celma, A.; Sancho, J. V.; Schymanski, E. L.; Fabregat-Safont, D.; Ibáñez, M.; Goshawk, J.; Barknowitz, G.; Hernández, F.; Bijlsma, L. Improving Target and Suspect Screening High-Resolution Mass Spectrometry Workflows in Environmental Analysis by Ion Mobility Separation. *Environ. Sci. Technol.* **2020**, *54*, 15120–15131.
- (30) Fabregat-Safont, D.; Ibáñez, M.; Bijlsma, L.; Hernández, F.; Waichman, A. V.; de Oliveira, R.; Rico, A. Wide-scope screening of pharmaceuticals, illicit drugs and their metabolites in the Amazon River. *Water Res.* **2021**, *200*, 117251.
- (31) Vera, P.; Canellas, E.; Barknowitz, G.; Goshawk, J.; Nerin, C. Ion-Mobility Quadrupole Time-of-Flight Mass Spectrometry: A Novel Technique Applied to Migration of Nonintentionally Added Substances from Polyethylene Films Intended for Use as Food Packaging. *Anal. Chem.* **2019**, *91*, 12741–12751.
- (32) D'Atri, V.; Causon, T.; Hernandez-Alba, O.; Mutabazi, A.; Veuthey, J.-L.; Cianferani, S.; Guillard, D. Adding a new separation dimension to MS and LC-MS: What is the utility of ion mobility spectrometry? *J. Sep. Sci.* **2018**, *41*, 20–67.
- (33) Paglia, G.; Astarita, G. Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry. *Nat. Protoc.* **2017**, *12*, 797–813.
- (34) Celma, A.; Ahrens, L.; Gago-Ferrero, P.; Hernández, F.; López, F.; Lundqvist, J.; Pitarch, E.; Sancho, J. V.; Wiberg, K.; Bijlsma, L. The relevant role of ion mobility separation in LC-HRMS based screening strategies for contaminants of emerging concern in the aquatic environment. *Chemosphere* **2021**, *280*, 130799.
- (35) Zhou, Z.; Luo, M.; Chen, X.; Yin, Y.; Xiong, X.; Wang, R.; Zhu, Z.-J. Ion mobility collision cross-section atlas for known and unknown metabolite annotation in untargeted metabolomics. *Nat. Commun.* **2020**, *11*, 4334.
- (36) McCullagh, M.; Pereira, C. A. M.; Yariwake, J. H. Use of ion mobility mass spectrometry to enhance cumulative analytical specificity and separation to profile 6- C /8- C- glycosylflavone critical isomer pairs and known-unknowns in medicinal plants. *Phytochem. Anal.* **2019**, *30*, 424–436.
- (37) Song, X.-C.; Canellas, E.; Dreolin, N.; Nerin, C.; Goshawk, J. Discovery and Characterization of Phenolic Compounds in Bearberry (*Arctostaphylos uva-ursi*) Leaves Using Liquid Chromatography-Ion Mobility-High-Resolution Mass Spectrometry. *J. Agric. Food Chem.* **2021**, *69*, 10856–10868.
- (38) Bijlsma, L.; Bade, R.; Celma, A.; Mullin, L.; Cleland, G.; Stead, S.; Hernandez, F.; Sancho, J. V. Prediction of Collision Cross-Section Values for Small Molecules: Application to Pesticide Residue Analysis. *Anal. Chem.* **2017**, *89*, 6583–6589.
- (39) Regueiro, J.; Negreira, N.; Berntssen, M. H. G. Ion-Mobility-Derived Collision Cross Section as an Additional Identification Point for Multiresidue Screening of Pesticides in Fish Feed. *Anal. Chem.* **2016**, *88*, 11169–11177.
- (40) Hines, K. M.; Ross, D. H.; Davidson, K. L.; Bush, M. F.; Xu, L. Large-Scale Structural Characterization of Drug and Drug-Like Compounds by High-Throughput Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2017**, *89*, 9023–9030.
- (41) Hernández-Mesa, M.; Le Bizec, B.; Monteau, F.; García-Campaña, A. M.; Dervilly-Pinel, G. Collision Cross Section (CCS) Database: An Additional Measure to Characterize Steroids. *Anal. Chem.* **2018**, *90*, 4616–4625.
- (42) Righetti, L.; Dreolin, N.; Celma, A.; McCullagh, M.; Barknowitz, G.; Sancho, J. V.; Dall'Asta, C. Travelling Wave Ion Mobility-Derived Collision Cross Section for Mycotoxins: Investigating Interlaboratory and Interplatform Reproducibility. *J. Agric. Food Chem.* **2020**, *68*, 10937–10943.
- (43) Song, X.-C.; Dreolin, N.; Damiani, T.; Canellas, E.; Nerin, C. Prediction of Collision Cross Section Values: Application to Non-Intentionally Added Substance Identification in Food Contact Materials. *J. Agric. Food Chem.* **2022**, *70*, 1272–1281.
- (44) Zhou, Z.; Shen, X.; Tu, J.; Zhu, Z.-J. Large-Scale Prediction of Collision Cross-Section Values for Metabolites in Ion Mobility-Mass Spectrometry. *Anal. Chem.* **2016**, *88*, 11084–11091.
- (45) Broeckling, C. D.; Yao, L.; Isaac, G.; Gioioso, M.; Ianchis, V.; Vissers, J. P. C. Application of Predicted Collisional Cross Section to Metabolome Databases to Probabilistically Describe the Current and Future Ion Mobility Mass Spectrometry. *J. Am. Soc. Mass Spectrom.* **2021**, *32*, 661–669.
- (46) Ross, D. H.; Cho, J. H.; Xu, L. Breaking Down Structural Diversity for Comprehensive Prediction of Ion-Neutral Collision Cross Sections. *Anal. Chem.* **2020**, *92*, 4548–4557.
- (47) Plante, P.-L.; Francovic-Fontaine, É.; May, J. C.; McLean, J. A.; Baker, E. S.; Laviolette, F.; Marchand, M.; Corbeil, J. Predicting Ion Mobility Collision Cross-Sections Using a Deep Neural Network: DeepCCS. *Anal. Chem.* **2019**, *91*, 5191–5199.
- (48) Gonzales, G. B.; Smaghe, G.; Coelus, S.; Adriaenssens, D.; De Winter, K.; Desmet, T.; Raes, K.; Van Camp, J. Collision cross section prediction of deprotonated phenolics in a travelling-wave ion mobility spectrometer using molecular descriptors and chemometrics. *Anal. Chim. Acta* **2016**, *924*, 68–76.
- (49) Mollerup, C. B.; Mardal, M.; Dalsgaard, P. W.; Linnet, K.; Barron, L. P. Prediction of collision cross section and retention time for broad scope screening in gradient reversed-phase liquid chromatography-ion mobility-high resolution accurate mass spectrometry. *J. Chromatogr. A* **2018**, *1542*, 82–88.
- (50) Groh, K. J.; Backhaus, T.; Carney-Almroth, B.; Geueke, B.; Inostroza, P. A.; Lennquist, A.; Leslie, H. A.; Maffini, M.; Slunge, D.; Trasande, L.; Warhurst, A. M.; Muncke, J. Overview of known plastic packaging-associated chemicals and their hazards. *Sci. Total Environ.* **2019**, *651*, 3253–3268.
- (51) Groh, K. J.; Geueke, B.; Martin, O.; Maffini, M.; Muncke, J. Overview of intentionally used food contact chemicals and their hazards. *Environ. Int.* **2021**, *150*, 106225.
- (52) Muncke, J. Reference Module in Food Science. *Chemical Migration from Food Packaging to Food*; Elsevier, 2016.
- (53) Hinnenkamp, V.; Klein, J.; Meckelmann, S. W.; Balsaa, P.; Schmidt, T. C.; Schmitz, O. J. Comparison of CCS Values

Determined by Traveling Wave Ion Mobility Mass Spectrometry and Drift Tube Ion Mobility Mass Spectrometry. *Anal. Chem.* **2018**, *90*, 12042–12050.

(54) Mullin, L.; Jobst, K.; DiLorenzo, R. A.; Plumb, R.; Reiner, E. J.; Yeung, L. W. Y.; Jogsten, I. E. Liquid chromatography-ion mobility-high resolution mass spectrometry for analysis of pollutants in indoor dust: Identification and predictive capabilities. *Anal. Chim. Acta* **2020**, *1125*, 29–40.

(55) Su, Q.-Z.; Vera, P.; Salafranca, J.; Nerín, C. Decontamination efficiencies of post-consumer high-density polyethylene milk bottles and prioritization of high concern volatile migrants. *Resour. Conserv. Recycl.* **2021**, *171*, 105640.

(56) Szöcs, E.; Stirling, T.; Scott, E. R.; Scharmüller, A.; Schäfer, R. B. webchem: An R Package to Retrieve Chemical Information from the Web. *J. Stat. Software* **2020**, *93*, 1–17.

(57) Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; Kuhn, M.; Pedersen, T.; Miller, E.; Bache, S.; Müller, K.; Ooms, J.; Robinson, D.; Seidel, D.; Spinu, V.; Takahashi, K.; Vaughan, D.; Wilke, C.; Woo, K.; Yutani, H. Welcome to the Tidyverse. *J. Open Source Software* **2019**, *4*, 1686.

(58) Djoumbou Feunang, Y.; Eisner, R.; Knox, C.; Chepelev, L.; Hastings, J.; Owen, G.; Fahy, E.; Steinbeck, C.; Subramanian, S.; Bolton, E.; Greiner, R.; Wishart, D. S. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J. Cheminf.* **2016**, *8*, 61.

(59) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V.; Tanchuk, V. Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl, P.; Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I. I.; Palyulin, V. A.; Radchenko, E. V.; Welsh, W. J.; Kholodovych, V.; Chekmarev, D.; Cherkasov, A.; Aires-de-Sousa, J.; Zhang, Q.-Y.; Bender, A.; Nigsch, F.; Patiny, L.; Williams, A.; Tkachenko, V.; Tetko, I. V. Online chemical modeling environment (OCHEM): web platform for data storage, model development and publishing of chemical information. *J. Comput. Aided Mol. Des.* **2011**, *25*, 533–554.

(60) Dong, J.; Cao, D.-S.; Miao, H.-Y.; Liu, S.; Deng, B.-C.; Yun, Y.-H.; Wang, N.-N.; Lu, A.-P.; Zeng, W.-B.; Chen, A. F. ChemDes: an integrated web-based platform for molecular descriptor and fingerprint computation. *J. Cheminf.* **2015**, *7*, 60.

(61) Elith, J.; Leathwick, J. R.; Hastie, T. A working guide to boosted regression trees. *J. Anim. Ecol.* **2008**, *77*, 802–813.

(62) Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery And Data Mining*; ACM, 2016; pp 785–794.

(63) Zhou, Z.; Tu, J.; Xiong, X.; Shen, X.; Zhu, Z.-J. LipidCCS: Prediction of Collision Cross-Section Values for Lipids with High Precision To Support Ion Mobility-Mass Spectrometry-Based Lipidomics. *Anal. Chem.* **2017**, *89*, 9559–9566.

(64) Righetti, L.; Bergmann, A.; Galaverna, G.; Rolfsson, O.; Paglia, G.; Dall'Asta, C. Ion mobility-derived collision cross section database: Application to mycotoxin analysis. *Anal. Chim. Acta* **2018**, *1014*, 50–57.

(65) Boschmans, J.; Jacobs, S.; Williams, J. P.; Palmer, M.; Richardson, K.; Giles, K.; Laphorn, C.; Herrebout, W. A.; Lemièrre, F.; Sobott, F. Combining density functional theory (DFT) and collision cross-section (CCS) calculations to analyze the gas-phase behaviour of small molecules and their protonation site isomers. *Analyst* **2016**, *141*, 4044–4054.

(66) Ubeda, S.; Aznar, M.; Nerin, C. Determination of oligomers in virgin and recycled polyethylene terephthalate (PET) samples by UPLC-MS-QTOF. *Anal. Bioanal. Chem.* **2018**, *410*, 2377–2384.

(67) Gallart-Ayala, H.; Moyano, E.; Galceran, M. T. Fast liquid chromatography-tandem mass spectrometry for the analysis of bisphenol A-diglycidyl ether, bisphenol F-diglycidyl ether and their derivatives in canned food and beverages. *J. Chromatogr. A* **2011**, *1218*, 1603–1610.

(68) Yang, Y.; Hu, C.; Zhong, H.; Chen, X.; Chen, R.; Yam, K. L. Effects of Ultraviolet (UV) on Degradation of Irgafos 168 and Migration of Its Degradation Products from Polypropylene Films. *J. Agric. Food Chem.* **2016**, *64*, 7866–7873.

(69) Hernández-Mesa, M.; D'Atri, V.; Barknowitz, G.; Fanel, M.; Pezzatti, J.; Dreolin, N.; Ropartz, D.; Monteau, F.; Vigneau, E.; Rudaz, S.; Stead, S.; Rogniaux, H.; Guillaume, D.; Dervilly, G.; Le Bizec, B. Interlaboratory and Interplatform Study of Steroids Collision Cross Section by Traveling Wave Ion Mobility Spectrometry. *Anal. Chem.* **2020**, *92*, 5013–5022.

## Recommended by ACS

### Machine Learning for Physicochemical Property Prediction of Complex Hydrocarbon Mixtures

Maarten R. Dobbelaere, Kevin M. Van Geem, *et al.*

APRIL 25, 2022

INDUSTRIAL & ENGINEERING CHEMISTRY RESEARCH

READ 

### Predicting Solute Descriptors for Organic Chemicals by a Deep Neural Network (DNN) Using Basic Chemical Structures and a Surrogate Metric

Kai Zhang and Huichun Zhang

JANUARY 07, 2022

ENVIRONMENTAL SCIENCE & TECHNOLOGY

READ 

### Global Historical Production, Use, In-Use Stocks, and Emissions of Short-, Medium-, and Long-Chain Chlorinated Paraffins

Chengkang Chen, Jianguo Liu, *et al.*

MAY 10, 2022

ENVIRONMENTAL SCIENCE & TECHNOLOGY

READ 

### MS2Tox Machine Learning Tool for Predicting the Ecotoxicity of Unidentified Chemicals in Water by Nontarget LC-HRMS

Pilleriin Peets, Anneli Kruve, *et al.*

OCTOBER 21, 2022

ENVIRONMENTAL SCIENCE & TECHNOLOGY

READ 

Get More Suggestions >